Article:
Manzano, A orcid.org/0000-0001-6277-3752 (2016) The craft of interviewing in realist evaluation. Evaluation, 22 (3). pp. 342-360. ISSN 1356-3890

https://doi.org/10.1177/1356389016638615

The Craft of Interviewing in Realist Evaluation

Introduction

Interviews are the most widely used method of data collection in social sciences. Benney's and Hughes (1956) famous quote 'sociology has become the science of the interview' is more relevant than ever. 21st century social researchers have become interviewers; interviews are their tools; and their 'work bears the mark' of those interviews (p. 137). Therefore, in evaluation research –most especially in formative, process and development approaches- interviews are frequently the main and only tool to generate data about programme effectiveness.

There is much debate about the conduct of the interview (closed versus open, formal versus informal, etc.) and there is also a rather elaborate embroidery of subtypes and distinctions. These are based on the interview processes (Oakley's (1981) feminist interview is a good example) and how the 'pseudo-conversation' is managed by the interviewer. Multiple similes are used to describe interviewer management skills: 'the maker of quilts' (Denzin and Lincoln, 2000); the dramaturgical 'performer' (Goffman, 1959), the 'portraitist' (Lawrence-Lightfoot, 2005) or the 'miner' of knowledge (Kvale, 1996). One contribution rather ignored - and a significant one for evaluation- is the realist interview. Realist evaluation proposes- among other methods of data collection- the use of theory-driven interviews to 'inspire/validate/falsify/ modify' (Pawson, 1996: 295) hypotheses about how programmes and interventions work.

It is beyond the scope of this paper to discuss diversity within realist and constructivist ontologies; (Clark, 2008; Madill, 2008) and the various subcategories of qualitative research, but it is important to recognise the myriad of approaches and combinations encapsulated in these terms (Maxwell, 2012). This paper aims to focus on the research act: the actual conduct of the qualitative interview within a realist evaluation research study, and it will focus exclusively on Pawson's approach to realist qualitative interviewing. The reader should note, additionally, that there are several approaches to realist evaluation (e.g., Henry, Julnes, and Mark, 1998) and to qualitative research (e.g., Hammersley, 1992, 2008; Maxwell, 2012, 2013). Therefore, there is no authoritative account of what the realist interview is, neither is there an authoritative version of the qualitative interview but this should not prevent us from exploring differences.

Maxwell (2012) noted that some realist studies retain an ontological realism while accepting a form of epistemological constructivism and relativism but, an explicitly realist approach to qualitative research is scarce (p. 6). Pawson and Tilley (1997) maintain that realist evaluators should interview differently from constructivist researchers; a new craft should be learnt to be able to conduct realist theory-driven interviews. Pawson and Manzano-Santaella, 2012 also claimed that a new species of 'qualitative realism' has emerged where descriptive qualitative accounts of stakeholders interpretations of both why a programme 'works' and the respondent's accounts of outcomes are interpreted as evidence of programme success. Following on from this view, the purpose of this paper is to add to the methodological guidance given by Pawson (1996) on how to perform realist interviews. As a point of reference, the paper starts by reviewing how interview techniques have been applied in published realist evaluations. Then, the paper suggests two guiding principles to interviewing when the researcher's programme theory is the matter of the interview (Pawson and Tilley, 1997). The first one relates to the design of realist studies (sampling strategy and choosing interviewees) and the second one explains how to ask questions like a realist, including the construction of topic guides and identification of three different phases in realist interviews: theory gleaning, theory refining and theory consolidation.

Qualitative Interviews in Realist Evaluations: Current Practice

In 1989, Pawson started exploring imposition and control in quantitative and qualitative research interviews (Pawson, 1989). In 1996, he published an article devoted exclusively to the interview (Pawson, 1996) and, immediately after, in the chapter 'How to construct realistic data: Utilizing stakeholder's knowledge' of the *Realistic Evaluation* book (Pawson and Tilley, 1997), the realist interview was explained and illustrated with examples. Theories are placed before the interviewee for them to comment on with a view to providing refinement. The subject matter of the interview is the researcher's theory and interviewees are there to confirm, falsify and basically, refine the theory. This relationship - described as a teacher-learner cycle - is distinctive of realist evaluations. It starts with teaching the interviewee 'the particular programme theory under test' and then 'the respondent, having learned the theory under test, is able to teach the evaluator about those components of a programme in a particularly informed way' (Pawson and Tilley, 2004, p.12). Therefore the roles of teacher and learner are not static but become interchangeable between the interviewer and the interviewee during the process of thinking through the complexities of the programme. Interviews used in realist evaluations are qualitative in terms of the conduct

because participant views are explored through conversations (and not ticking boxes) but the purpose of the interview is the key difference. Maxwell (2012) explains how in realist studies data collected through qualitative means are not considered constructions. Data are instead considered "evidence for real phenomena and processes" (p.103). These data should be used to make inferences about such phenomena and processes, and therefore re-testing them against additional/alternative data should be an inherent part of the research practice.

Many useful handbooks, textbooks, journals and web resources exist on how to conduct qualitative interviews (Gubrium, 2012; Mason, 2002; Holstein and Gubrium; 1997, Mishler, 1991; Briggs, 1986). Technically, just in terms of how questions are asked, qualitative interviews range from open ended conversations to using semi-structured or structured prompts and guides. In these conversations researchers tend to explore aspects and concepts, while realist interviews are interested in investigating propositions. Although most qualitative researchers also use theory in planning and assessing their interviews, Pawson proposes that theory should be used more explicitly and systematically. There is, of course, a continuum, in theory-driven qualitative approaches to data collection and it is not possible to establish a black-and-white distinction between them. In his article *Theorizing the Interview*, Pawson (1996) criticises the simple division of interviews into structured or unstructured advocating for theory-driven interviews based on the 'learner-teacher cycle'. The conventional technical guidance is shifted and the interviewer needs to abandon the traditional neutral territory to engage with respondents. The evaluator/interviewer does not take an insider or an outsider perspective about the programme and this is not an easy manoeuvre.

To understand the current state of qualitative realist research interviewing, the research methods section of realist evaluation studies in the health sector published between years 2004-2013 inclusively were examined. The review was focused around three main questions: (1) What are the research methods used in these realist evaluation studies? (2) If interviews were used, did they apply the learner-teacher cycle? (3) How was the cycle applied and what methodological problems related to interviewing do these papers raise? In 2012, Marchal et al. reviewed realist evaluations following Pawson and Tilley's approach published in applied health studies between 2004 and 2010. Three more years of published realist evaluations in health were added to their list and a total of 40 papers were found; of those, 32 papers included interviews as one of the methods of data collection (See Figure 1). Five papers used interviews as their only method of investigation and the rest (n=27) employed interviews combined with other methods. The most common pattern was to combine mainly multiple qualitative methods: semi-structured interviews with documentary

analysis (sometimes this included secondary analysis of quantitative data) and ethnographic observation. However, only 8 of the 40 papers mixed quantitative and qualitative primary data.

This brief review demonstrates that qualitative interviewing is the most common method of data collection in realist evaluations and it is normally used in combination with other methods, which (in the published studies in health) tended to be also qualitative. Surprisingly, qualitative interviewing is commonly treated as unproblematic with little attention given to fieldwork processes or the act of the interview itself. Most articles include a brief mention of how many interviews were conducted, and who the interviewees were plus verbatim respondent quotations. Much of the data collection appears to use traditional semi-structured approaches asking about programme effectiveness and barriers with little explicit reference to the delivery of interview questions that can demonstrate how data are collected with the aim to make inferences about the phenomena and processes evaluated. The use of standardised interview guides is also frequently reported with some brief explanation (topic guides are not included in the papers) given of the content of the questions – i.e. they are generated within in a theoretical framework, which is the case with all interview methods. For example: 'we used an interview guide derived from the main concepts in our theoretical framework' (Etheridge et al., 2013, p.5); and 'An interview guide was developed and consisted of a series of questions asking respondents to describe factors that facilitated or impeded the adoption and implementation of [the programme]. Respondents were also asked to identify changes in the planning and priority-setting process over the previous two years. Consistent with qualitative research methods, an open stance was maintained, probing into emerging themes and seeking clarification when necessary' (Maluka et al., 2011, p. 5). These descriptions are analogous to those used in non-realist studies making it difficult to assess whether they are consistent with teacher-learner interviews. A discussion of the theories carried on in questioning and the contextualisation of those questions would have been helpful for the reader.

Methodological problems were frequently raised by researchers applying the realist approach (Marchal et al., 2012) and expressed mainly as analysis difficulties to distinguish between contexts (C), mechanisms (M) and outcomes (O) but none questioned the application of the techniques to ascertain those. Although word limitations frequently constrain the reporting of technical troubles, perceptions of 'illusory' simplicity (Kvale and Brinkmann, 2009) also apply in other interviewing approaches which are often treated as unproblematic. There seems to be a need for technical guidance on how to conduct realist interviews but this need is not always explicit. Anecdotal evidence from encounters with practitioners contradicts the 'unproblematic' hypothesis. Evaluators do find the prospect of

realist interviewing challenging and common questions asked by novice colleagues to other realist evaluators are: 'How do you prepare realist topic guides?'; 'How do you present topic guides for realist evaluations to ethics committees?'; 'How do I do iterative analysis when I am only doing one lot of interviews?'; ' How do I talk to interviewees about Cs, Ms, Os?'; 'I have prepared a table with hypothesised CMOs, should I show it to the participants?'; 'Can I use NVIVO to code CMOs from realist interview transcripts?' Anxieties seem to be high before entering the field and before starting the coding of transcripts.

Considering all of these questions, the second part of this paper, proposes two main principles aimed to provide some guidance - a general set of criteria rather than rigid rules - to enhance the interviewing skills of realist evaluators before, during and after leaving the field. The first and main principle contextualises the interview in scientific realism, the philosophy of science framing the realist evaluation approach. This ontological position influences epistemology and design choices (respondents, sampling, etc.). The second guiding principle is more practical advice on how to conduct the interview. Finally, a brief reminder of the iterative principle in realist data analysis is given.

Guiding principle 1: Before the interview. Implications of the philosophy for design

In the last two decades, commissioners around the world have started to fund realist studies and this is now a well-established approach rapidly growing in popularity in several disciplines but mainly in applied health research, an area traditionally dominated by polarised methodological battles (Greenhalgh, 1999). While realist methods seem to have been embraced around the world, it is not always clear which is the ontological basis of self-named realist evaluation studies. It is important to remember that the philosophical basis of realist evaluations and realist reviews is realism, which assumes an external reality that can be assessed through configurations of contexts, mechanisms and outcomes. The premise is that volition, change and conditionality must be assumed and pursued in realist evaluations of social programmes (Pawson, 2013). Furthermore, theoretical claims obtained through these evaluations are expected to represent knowledge of the real world.

A qualitative approach can help in the identification of contextually-grounded explanatory mechanisms which could be difficult to obtain using only quantitative methods (Sayer, 1992). It is not always clear that the realist-reorientation of the qualitative enquiry (Iosifides, 2011) has permeated the data collection process, even when there is clear evidence of an extensive use of qualitative interviews for causal, explanatory purposes in realist evaluation studies. Ideally, interviews should be used as means to explore propositions that will be tested and refined with other data and not just as a means to an end. Hammersley (2008, p.30) explained that relying primarily or exclusively on qualitative interviews 'suffers from

some of the same problems as reliance on attitude inventories in survey research. In many cases, it assumes that behaviour is in large part a function of some stable orientation that directly expresses itself in the same way in diverse contexts'. Evaluation research, however, takes place within a political, organisational and financial context, and even when design choices are informed by theoretical considerations, they are also framed by practical difficulties (i.e. the availability of various forms of data, feasibility of collecting data within financial, time and access constraints).

The realist premise is that the research process will start by theorising, then will test those theories, these will be refined and tested again and, in this iterative process, our understanding of the real world is also refined. This logic has significant implications for the research design. The choice of data collection methods should be theory-driven and they should be purposely selected because these specific methods are supposed to be the best tools to uncover patterns and regularities about the programme. For example, in complex programmes random sampling may not be able to identify the variation of programme participants because the intervention has unexpected processes that cannot be predicted *a priori* for statistical purposes. Interviews may help exploring theories that try to understand those unexpected processes. Once qualitative interviews have been chosen because they seem the best available and feasible tool for the evaluator to explore specific theories, there are still a few other 'realist' decisions to be made like choosing with whom, how and how many times we are going to interview the same/different people. These are discussed in the next two sub-sections:

A) *How many interviews and how many set of interviews? Sampling strategy in realist evaluation studies*

Traditionally qualitative enquiry advocates for a small number of interviews and although the precise number cannot be decided *a priori*, samples are normally smaller than in quantitative research. Calculations are not based on notions of sampling fractions, representativeness and bias elimination but on the concepts of data completeness and saturation (Corbin and Strauss, 2008; Guest et al., 2006). There is widespread agreement among qualitative methodologists that there is no set number of interviews that can be assumed to achieve saturation (Morse, 1995; Saumore and Given, 2008). There are also, however, over one hundred sets of proposals about 'quality' criteria in qualitative research, featuring several non-reconcilable positions (Dixon-Woods et al., 2004). Pragmatic informal rules permeate and often lead research proposals when numbers must be stated a priori in front of funders, commissioners or ethics committees. Despite the conceptual impossibility of establishing a

power calculation, common professional practise situates the acceptable number of interviews between 20 and 30 (Mason, 2010)]. This rule of thumb is not strictly reliable for realist evaluation studies. The reason for this is that realist hypotheses are not confirmed or abandoned through saturation obtained in a double-figure number of qualitative interviews but through relevance and rigour (Pawson, 2013) obtained in a mixed-method strategy. A theory may be gleaned, refined or consolidated not necessarily in the next interview, but also while digging for nuggets of evidence (Pawson, 2006) in other sources of data (i.e. documents, routinely collected administrative data). Emmel (2013) has written extensively about the principles of realist sampling and how to deal with the 'allure of the number n' (p.146). Several of his key points are relevant when embarking on realist evaluation interviews and are explored below:

1) 'All we have are fragments' (Emmel, 2013, p.141) and those fragments should be explored in-depth by the evaluator. Emmel (2013) explains that realists try to understand how each fragment of evidence contributes to their interpretations and explanations; and how their ideas are tested and refined within and between those fragments. This principle reminds the researcher of the need for other methods of investigation to build up theories from relevance to rigour (Pawson, 2013).This multi-method approach is not only a triangulation friendly tactic useful at the analysis stage, it embeds the realist fieldwork experience. In addition, one of the purposes of interviews is also to build knowledge of variations in what happens in natural settings and this knowledge contributes to building, testing and refining theories. The realist researcher-whenever feasible- should arrive at the interviews knowledgeable of what happens in the natural setting. This could be because for example, previously staff meetings have been observed or because the grey literature for the programme has been studied in detail. Methods are chosen to complement each other and to fill each other gaps (Patton, 2002). Theories gleaned or further developed in the interviews are refined against the observational and documentary data and vice versa, all methods having a 'conceptual refinement function' (Pawson and Tilley, 1997, p. 167).
2) The proposed number of interviews to be carried out mentioned in research proposals is only an approximate plan because the realist process of theory-testing is unpredictable, unstable and uncertain. This, however, can be problematic in contracted evaluations, which are often short term, under-funded, and very precisely costed . Emmel explains that the essence of realism implies that 'in realist qualitative research the sample can only be weakly elaborated beforehand' (Emmel, 2013, p. 154). A rough idea of sample size can be clarified soon after fieldwork commences.

Purposive sampling can be based on different criteria and of these, the need to 'maximize variability so as to discover whether the program succeeds across a whole spectrum of sites' (Weiss, 1998, p. 164) should drive the sampling of participants. Realist sampling should be designed to test the contexts that are hypothesised to matter. This might not be only about sites. It might be also be about population groups, or about implementation barriers and facilitators. Evaluators become knowledgeable of programme successes and barriers as soon as they start conversing with front-line staff and it is at that point that theories start to develop shape and the approximate number of cases the interviewer will seek to pursue can be established.

3) Regardless of the final number of interviews, realist enquiries should aim to collect large amounts of data. Substantial amounts of primary or secondary data are needed -even when the sample is small to move from constructions to explanation of causal mechanisms. Since the unit of analysis is not the person, but the events and processes around them, every unique programme participant uncovers a collection of micro events and processes, each of which can be explored in multiple ways to test theories. Greenhalgh et al (2009) in their realist evaluation of a large-scale healthcare intervention in London conducted 100 interviews and these were also accompanied by observations, group interviews, informal discussions, documentary analysis and secondary analysis of quantitative and qualitative documentation. Rycroft-Malone et al (2008) combined 'non-participant observation of nursing and multi-disciplinary activities; semi-structured interviews with practitioners and patients involved in observations; semi-structured interviews with other stakeholders; tracking patient journeys; field notes and the review of relevant documentation (e.g. copies of pathways, guidelines)'. In (Manzano-Santaella, 2011) realist evaluation of hospital processes although 'only' 14 case studies were studied, a total of 39 semi-interviews, 73 observations and 93 other activities (phone calls, conversational interviews, informal observations, etc.) took place.
4) To be able to build explanations an iterative process of data collection should be designed which includes the possibility to revisit respondents and repeat interviews (formally or informally) with the same participants (or sub-groups) at a later stage of the investigation. This longitudinal element is different from standard longitudinal qualitative research because, although it involves returning to interviewees, the objective is not to explore changes which occur over time in participants' lives but to explore and further develop evaluator's theories as he/she is becoming more knowledgeable about the programme. Changes have occurred in the evaluators theories and these newly refined theories are passed by those who know the

programme well in search for further refinement. Repeated interviews can also be scheduled after observations and then these are guided and informed by incidents arising from observations.

Repeated interviews with the same respondents are not always a feasible option but most importantly, this is not the only strategy to develop theories. One can do it by asking different or additional questions in future interviews with different respondents, by going back to the literature, by observing key processes, etc. There are several constraints to this flexible approach (resources, ethic approvals, and logistics) but as much as possible this flexibility should be incorporated in those 'weakly elaborated plans' (Emmel, 2013). In summary, the importance is not on the 'how many' people we talk to but on the 'who', the 'why' and the 'how'. We are trying to understand how our interviewees understand and have experienced the programme and compare those experiences with our hypotheses about how the programme is working.

B) From whom. Choosing and finding knowledgeable interviewees

Pawson & Tilley (1997) recommended respondent selection to be based on their 'CMO investigation potential'. Each component, contexts and mechanisms and outcomes, triggers the need for a different kind of respondent. Their view is that practitioners have specific ideas on what is within the programme that works (mechanisms) because they are likely to have broad experience of successes and failures, and some awareness of people and places for whom and in which the programme works. These individuals tend to have more quality information regarding the underlying programme theory. The imperative is to work with a broad range of programme stakeholders who must be purposively selected based on evaluator's hypotheses. Evaluators carry embryonic or well developed theories into the encounter with programmes because they have knowledge from similar and previous programmes and knowledge from social science theory (Pawson and Tilley, 1997; Chen and Rossi, 1992).

For initial theory gleaning, it is better to start interviewing practitioners than users of the programme: people who know the programme well, for example those whose job it is to monitor what goes on (i.e. middle managers, ward managers, etc.). Manzano & Pawson (2014) organised several events with organ donation coordinators from Spain and the UK before initiating a comparative evaluation of deceased organ donation systems. The tacit knowledge of these practitioners helped in interrogating in closer detail the programme

theory underlying the donation to transplantation processes found in the literature before entering the field. Those conversations revealed how steps in the organ donation process were negotiated locally and how practitioner tacit knowledge re-shaped activities in each donation system. CMOs started to take shape in those conversations, ready to be tested in future ethnographic observations and interviews with other stakeholders.

Frontline practitioners could be the next set of interviewees because they frequently see themselves as 'picking up the pieces' following top-down programme implementations and are good sources of information about programme barriers and unintended consequences. It must be assumed that different practitioners will experience the programme processes differently so their varied experiences need to be captured. Different points of views must be pursued not to ensure balance or consensus (Blamey and Mackenzie, 2007) but because a variety of perspectives are needed to investigate informal patterns and unintended outcomes. The first set of interviewees may mention other people, other institutions, as they describe cases and events to illustrate programme dynamics. If they do not suggest other stakeholders, the evaluator can ask them to do this. Letting interviewees know that others are interviewed with opposing or different views, may influence participants' responses offering more detailed explanations (Rubin and Rubin, 1995, p. 70). Knowing that the evaluator may enquire about the same phenomena to other staff who may have competing perspectives can be used as an strategy for encouraging interviewees to think about the programme in general further from their specifics of their own experiences.

Subjects of the programme (i.e. service users, patients, claimants) are more likely to be sensitised about outcomes. They are not great 'mechanism experts' in the general sense, rather they are experts on how some of the programme mechanisms may have influenced some of their outcomes (Pawson and Tilley, 1997). It must be noted that sampling strategy here is more complicated because information-rich cases (Patton, 2002) will be difficult to guess when our knowledge is still tentative. For example, in a study evaluating a programme to accelerate hospital discharges (Manzano-Santaella, 2011), several formal and informal meetings took place with stakeholders (i.e. ward managers, discharge liaison officers) to discuss how and where to recruit patients with different illnesses that could represent different challenges for the theories to be tested and refined. Consequently, face to face interviews with patients allowed the researcher to cover a different range of discharge experiences in 14 very diverse patients ranging from a 40 year old man who needed re-housing after a stroke left him unable to use the stairs in his flat; to an 87 year old woman with multiple long-term pain conditions who was admitted into hospital due to breathing

difficulties caused by a chronic respiratory illness. Although sampling was informed by theoretical considerations (i.e. testing theories about reasons why some people stay in hospital longer than other), it was, however, also framed by practical difficulties. Numerous possible participants were lost in medical 'contingencies' and the social organisation of medical work (Strauss et al, 1985): people were transferred to other wards, to intensive care units, discharged without previous notice or they died.

Guiding principle 2: Ask questions like a realist

It is important to remember that despite using much of the paraphernalia of qualitative inquiry the objective of a realist interview is not to elicit participant narratives. The realist interview is conducted within a programme evaluation context. It is the programme's story that we are pursuing and we will do it by capturing the participants' stories because those experiences of the programme illuminate the varying processes and manifold outcomes of the programme (Patton, 2003). Rubin & Rubin (1995, p. 20) differentiate between 'cultural interviews' which explore norms and values and 'topical interviews' –like the interview to evaluate programmes- which focus on events or processes. The ability to trace processes requires that the researcher takes an active role in directing the questioning and keeping the conversation on the specific topic under evaluation. Additionally, it must be noted that the term 'process evaluation' is contested and embedded with different meanings (Linnan and Steckler, 2002) but it is normally equated with measuring the extent to which interventions are implemented and delivered as planned. Tracing processes of implementation is a frequent part of realist evaluations because stakeholders' meanings and reasoning processes about programme implementation can help identify key contextual differences in the construction of the outcome patterns.

The traditional advice on how to conduct qualitative interviewers is to pretend incompetency to avoid data contamination and to act as a *deliberate naiveté* (Kvale and Brinkmann, 2009). In this aspect, some guides to qualitative interviewing do not differ from those oriented to standardised surveys – with primary concerns being about maximizing the flow of valid information through minimising distortions. This quote from a manual addressed at healthcare researchers is very clear on the concealing technique: 'You will want to conceal your knowledge as far as possible. The correct role for the qualitative researcher has often been described as that of the 'amiable incompetent' – someone friendly and intelligent but lacking knowledge, someone who has to be *told* things' (Sapsford and Abbott, 1992, p. 112). Another common recommendation is 'to be neutral towards the topic while displaying interest' (Fielding and Thomas, 2001, p 129). Much of the methodological literature on interviewing cautions interviewers to be careful in how they ask questions aiming to obtain

'unadulterated facts and details, most of which rely upon interviewer and question neutrality' (Holstein and Gubrium, 1997, p. 116).

This prevailing textbook advice followed by many novices is not useful for conducting realist interviews because in these type of interviews, the interviewee is not supposed to control the direction of the conversation, it is the interviewer who will aim to take control and to steer the interview towards his/her topic of interest. Experienced evaluators control the direction of conversation in most interviews but in realist evaluations they are not only asking about the programme but about the theories that the programme assumes. This relationship is contextually different from ethnographic interviews because it occurs in an evaluation context and evaluation, for realists, is 'assisted sensemaking' (Mark et al., 1999). The relationship with the interviewer has been deliberately and artificially created for interrogating ideas about the programme. It is no natural relationship; it is in fact 'special, artificial and somehow disembedded' (Dahler-Larsen, 2011). This control must be reflected when preparing the instruments for collecting the data (topic guides) and also in the field, during the interviews.

1. *From mind to paper: Scripting for realist interviews*

One could say that Pawson and Tilley's work is an on-going quest for the right interrogative pronouns when re-phrasing the evaluative research question. Tilley in his chapter 'What is the what in what works' (Tilley, 2009) identifies five possible meanings of the interrogative pronoun 'what' in the fields of health and crime prevention. These are: particular intervention, class of measures, mechanism, strategy or outcome patterns. This clarification illustrates the importance of choice and phrasing of questions because one of the points of the realist interview is to recount experiences and reasoning related to their context (Campbell, 1975) and to the emerging theories that are under investigation.

Realist interviews are generally semi-structured, containing exploratory questions based on the programme evaluated but acting as instruments to draw out the propositions of the general inquiry. It is also possible to begin with more structured questions as long as respondents are given space to explain their initial responses. The topics covered should reflect the objectives of the study as a whole, and be concerned with the provision of information around services, specific arrangements and resources for specific users or teams, and other relevant resource issues. Exploring differences in implementation might trigger hypotheses about different mechanisms and outcomes. When interviewing different people about different aspects of the programme theory, interviews should take different foci. Interviews should be designed around stakeholders' awareness and experiences of the programme, including their reasoning (Dalkin et al, 2015) about specific propositions.

Some of the conventional guidance on how to design scripting instruments for qualitative interviews (interview protocols or topic guides) aims to ensure interviewer standardisation and consistency (Boyce and Neale, 2006). Those trained in more positivist methods are also likely to assume that testing hypotheses requires asking the same/similar questions to everyone even if these are open-ended. The realist mantra, however, is based on the notion that 'nothing works unconditionally in all circumstances' (Tilley, 2000, p. 126) and this notion has a direct and radical effect on the evaluation questions. Intra-contextual variation impacts on the questions which should be qualified by circumstances, temporality and sub-groups. Homogeneity contradicts realist thinking as the premise is that knowledge will evolve and questions will change as answers alter evaluators' knowledge. Although realist and constructivists use non-standardised approaches to hypothesis testing, there is a significant difference between realist and constructivist prescription on how to ask questions, exemplified in Seidman's advice: 'In-depth interviewing, however, is not designed to test hypotheses, gather answers to questions, or corroborate opinions. Rather, it is designed to ask participants to reconstruct their experience and to explore their meaning' (Seidman, 2012, p.94). Conversely, realist interviews are, in fact, designed to test hypotheses. Programme theories are the subject matter of the interview and these are hypotheses that need to be elicited, developed, refined and tested. The realist topic guides reproduced in Table 1 and Table 2 illustrate how the evaluator's pursuit of precise knowledge is behind each question. Questions are constructed to test investigator's hypotheses and the stories of the interviewee will help refined them or discard them. In the first interviews, those hypotheses start in the form of possible Cs, possible Ms and possible intended and unintended Os. As more interviews are conducted and more evidence is gathered from other sources (i.e. observations, policy documents, other evaluation of the same programme, surveys), interviewer hypotheses will aim to construct outcome patterns (CMO). The notion of emergence is another tenet of realist evaluation (Pawson, 2013) and interviewers should also plan for the unplanned and be ready for the exploration of unexpected (not previously hypothesised) Cs, Ms, Os.

The overall purpose of the realist interview is distinctive, and its exact form will differ according to the 'ground gained' in the investigation. The process of elucidating and refining hypotheses is iterative, and therefore the context in which each of the realist interviews occur is dissimilar. This has a direct consequence to the design of realist topic guides which should reflect those dynamic principles and those interview contexts which are explored in the next section.

2. *Programme theory and phases in realist interviews*

For the purpose of this explanation and taking into consideration the evolving nature of the realist interviewer knowledge (learner/collaborator/teacher), three phases of realist interviews are proposed: theory gleaning, refinement and consolidation. All research designs have sequential phases (exploratory, data collection phase, triangulation and further analysis phase) and these can be used to locate and coordinate the interview through the phases of the research cycle. They illustrate how the evaluator adjusts and shapes the interview, keeping theory as the common denominator. Sometimes these three phases will happen within the same evaluation but due to constraints, some evaluations may have to stop at phase one or two. Other studies may already know enough about some theories and they start at phases two or three:

Phase 1. Theory gleaning interviews. In the first set of interviews, respondents are expected to help the evaluator to articulate first order theories which are those that identify how the contextual circumstances of some of users/programmes may impact behaviour and effectiveness. Research proposals of realist evaluations should ideally contain tentative theories of the middle range about how the programme is supposed to work and some of its assumptions will be tentatively articulated through a range of strategies (i.e. theory-driven literature review, expert panels). As it can be seen in the example of 'theory gleaning' topic guides shown in Tables 1 and 2, in practice, questions should start ascertaining which of the hypothesised contexts, mechanisms and outcomes the stakeholder will be able to comment on. First at their meso-level (because of their role in the process) and then at the micro-level (because of who they are e.g. more experienced, more well connected). That is, interviews should start with general questions about interviewee role/experiences/views about the programme and then following up on what they say by asking them to tell their stories (Seidman, 2012) about specific experiences or issues with programme participants/constraints. The questions asked in these first set of interviews will be mainly exploratory and the wording of those questions should try to ascertain how the programme works for whom and in what circumstances. Questions looking to explore context may ask for interviewees' experiences of before/during/after the programme was implemented. For example 'How was your work different before the programme was implemented?' .' Is this new programme going to work for everyone?' 'Could you explain to me the types of people and places where you think it may be more effective?' Stronger questions about context

should encourage people to compare subgroups, location, times, before and after. The objective is to draw the interviewee into comparison to explore contextual effectiveness. Stake (2010) explains that one of the purposes of qualitative interview is 'finding out about 'a thing' that the researchers were unable to observe themselves' p.95. With those examples, events, stories and cases, the evaluator eventually will be able to glean tentative explanations and from there to look for other potential interviewees, observations, comparisons and so on to build rigour from those relevant (Pawson, 2013) stories.

Phase 2. Theory refinement interviews. In these follow-up interviews, second level theories are incorporated in the evaluators thinking process. Now, the evaluator is becoming more knowledgeable of programme nuances, and questions evolve to being less standardised and more tailor-made to refine specific outcome patterns. Meanings attributed by the evaluator to previous answers or situations observed are discussed in the light of potential divergence and they are presented to the interviewee while spelling out the evidence (i.e. 'When they implemented this programme in London, they had this problem with...Have you seen that in this locality? If no, why do you think this is the case? How are the two localities different?'; 'The guidance about this programme says that this is supposed to happen, but when I talked to your colleague (or when I observed that meeting), I noticed that this was not the case. Have you experienced this yourself with any of your users? If no, why do you think this is the case? If the answer is yes, could you tell me the story of that user'). The following extract from a realist interview (Manzano-Santaella, 2011) illustrates how individual cases can be used as prompts to refine general programme theories while exploring unobservable events or thought processes. In this example, the evaluator was trying to explore mechanisms leading to people being transferred into care homes to accelerate hospital discharges. The interviewee refined the evaluator theory, adding that some patients consider care homes a safer option than returning home, especially after a long period of acute illness.

Evaluator: Why do you think this patient was discharged to a nursing home? Do you think he could have gone to his own home instead of a nursing home? And I am saying this because one of the theories about this policy is that to accelerate hospital discharges, is sending people into care homes too soon. Right?

Discharge liaison nurse: Ummm…I think at the time, this patient could have gone home and managed at home with a big care package. We could have organised three to four home care visits a day- and one visit in the night time. I think, he has gone into care because he'd been in hospital for a long time and he was scared about being on his

own. Plus he had some medical issues which needed monitoring. And I think, possibly, it was also peace of mind.

As the example showed, tentative theories (ideas) about how the programme works are spelled out and exposed to the interviewees, using their expertise as tools for hypotheses refinement. Then, as the number of cases cumulates, comparisons are made with all participants and with data generated by observations. In the realist evaluation approach, although stakeholders are not necessarily members of the evaluation team, they always play an important role in the development of theories. Not as mere 'researcher's subjects' (Patton, 1999) but as key informants with the power of their knowledge about how the programme is really operating.

Phase 3. Theory consolidation interviews. These three phases are interlinked, and this third phase also relates to hypotheses refinement. In the preceding phase, the evaluator selected a number of candidate theories, and some of those theories were discounted. In this third phase, the theories more worthy of consideration are fine tuned. This is a second level of refinement because, theory refinement is a never-ending task that will follow on even after our study is finished. In these final interviews, third level theories are established. The consolidation process requires the investigation of other existing rules, protocols and unwritten norms about programme users. Theories at this level should look for how new interventions or programmes modify the way that routine roles and behaviours are enacted. The 'modification' could involve increased attention, less attention, or business as usual. These interviews aim to further refine how this emerging theory performs for different configurations of stakeholders (patient, nurse, doctor, visiting relative, etc.) and the stakeholder response (relax, maintain vigilance, more specified positive action, etc.). Conversations with stakeholders should be guided with the help of the specificities of the individual cases, and from there, they can be directed into the general programme. In summary, this phase gives more detailed consideration to a smaller number of CMOs which belong to many families of CMOs. The role reversal feature (Pawson & Tilley, 1997) which is also present in the previous phases, consists here in the evaluator presenting his now nearly consolidated theory to the interviewer based on his/her own stories or nuggets of evidence (Pawson, 2006). The interviewer may help this consolidation with further insights, facilitating the interplay between tentative theories (conjectures) and error elimination (refutation) with their own validating or falsifying examples (Popper, 2002). In Box 1, an example is provided of how realist evaluators become knowledgeable through their fieldwork stories, in this case

the unintended outcome of how patients are quickly discharged from hospital without involving them in detailed planning.

Finally, a disclaimer must be noticed. The aim of identifying these three phases is to better understand how interviewer knowledge evolves and to assist evaluators in data collection processes and not a prescription or a phase- labelling exercise that could constrain interviewers. Although different emphases occur at different stages and these vary according to the level of theory refinement, it must be noted that three rounds of interviews are not always necessary to progress through these phases. For example, experienced evaluators, can progress through phases one and two within the same interview; or the third phase can be consolidated with another type of data collection like observations or audit data.

Reminder: The iterative process of realist analysis. Refining programme theory to and from the evidence

Guidance on how to analyse data in realist evaluations is abundant (Dalkin et al., 2015; Pawson, 2013; Westhorp, 2013; Pawson and Tilley, 1997). However, the realist analysis is not a technical process consisting of coding verbatim text with the help of a computer software package once fieldwork has been completed, then trudging out a few themes constructed from multiple subthemes and labelling them as contexts, mechanisms and outcomes. Realist analysis is not a defined separate stage of the research process; it is an ongoing iterative process of placing nuggets of information (Pawson, 2006) within a wider configurational explanation (CMO). Analysis should start ahead of data collection and insights pursued should ideally be both contemporary with and retrospective to fieldwork. Therefore the analysis of realist datasets is also constructed and timed differently.

Conclusion

Craft makers and fine artists learn by working in the apprentice system. All art is crafted although not all craft is art. In open-ended interviews the researcher is the 'prime research instrument' (Yin, 2010). Like craft makers use their hands, evaluators use interviews to collect and construct data. The craft of interviewing in realist evaluation studies is an important part of developing knowledge and it should be taught and learnt. Sharing scholarly concern that ontological and empirical principles are not always followed in realist evaluation studies, this paper has intended to make a modest, practical contribution to this end. A review of current interview practice demonstrated that although the qualitative interview is

the most commonly used research method in realist evaluation, the interview process seems to be treated as rather unproblematic and the realist-reorientation of the qualitative enquiry is not always evident. The main guiding principle explained is that realist evaluations are not constructivist studies despite using qualitative enquiry and consequently, design and fieldwork activities should aim to theorise, then test those theories, these will be refined and tested again and, in this iterative process, our understanding of the real world is also refined.

Acknowledgement

I would like to acknowledge Prof Ray Pawson, Dr Nick Emmel and Dr Gill Westhorp for their methodological guidance and expertise throughout the refinement of this article. Useful comments were also received from two anonymous referees.

Tables & Figures

Figure 1: Interviews in realist evaluations. Current practice

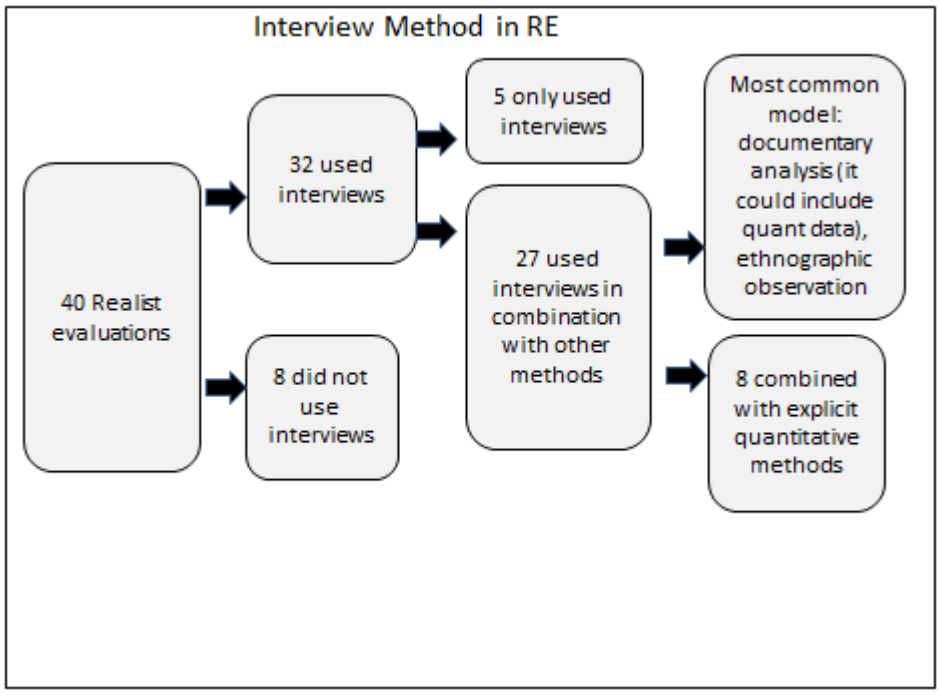

Table 1- Topic guide for qualitative interviews with hospital staff in charge of patient 1 discharge

	QUESTION	LOGIC
1	Could you explain your reasoning when organising this patient's discharge?	Questions 1-3 are introductory, to get them talking.

2	What ideally should happen to this patient in terms of discharge?	Questions 1-3 are introductory, to get them talking.
3	Could you talk to me about any difficulties that can alter your plans?	Questions 1-3 are introductory, to get them talking.
4	What characteristics in this patient would suggest that they may be likely to get delayed? I am thinking age, mental health condition, finances …	EXPLORING Context 1 (patient characteristics)
5	What characteristics in the way the staff works with this patient will help create a faster discharge?	Questions 5 & 6: Exploring Context 2 (Staff characteristics)
6	How do you think the level of experience in discharge planning in the members of the team has influenced the way this case has developed?	Testing Contex2-staff experience. Questions 6-9. Looking also for mechanisms leading to people being transferred out of hospital faster. Asking about other staff first and then about themselves.
7	There seem to be external factors affecting the way this patient's discharge plan progresses, I am not talking now about family but more about things like how the bed situation (lack of) in the ward may influence some of the decisions made…	Exploring Context 4- Characteristics of the infrastructure. Questions 6-9. Looking also for mechanisms leading to people being transferred out of hospital to accelerate hospital discharges. Asking about others first and then about themselves.
8	How do you think the new fine system has impacted how social services staff dealt with this case? I am thinking that they may be doing things differently than they used to do before the new programme was implemented?	Questions 6-9. Looking for mechanisms .Asking about other staff first and then asking about themselves.
9	How do you think the new programme has affected the way the multi-disciplinary team works as a team?	Question 6-9. Looking for mechanisms. Asking about other staff first and then about themselves.
10	Now, let me ask you about any changes you see in your own professional practice because of the new	Looking for Unintended Outcome 1 (Patients are discharged without

	programme? Prompt: I am thinking about the practicalities of involving this patient in the discharge process when the time is so limited.	meaningful involvement in decision making)
11	Are there any important things that are working well for your professional practice with this patient because of the new programme?	Looking for known, unknown intended and unintended outcomes
12	I just wondering whether the new programme may create higher stress levels than usual and how that has affected the relationship with the social services colleagues in this case?	Looking for Unintended Outcome 3 (Relationships and motivation that used to be based on trust are transformed by the financial deterrent)
13	In your opinion how appropriate is the resource (e.g. nursing home, home care) used to discharge this patient sooner? If negative or ambiguous answer: In your view, what would it be an appropriate outcome for this patient?	Looking for known and unknown outcomes

Table 2. Interview topic guide for social services staff

	QUESTION	**LOGIC**
1-11	Questions 1 to 11 are the same as in Table 1	
12	Suppose the hospital referred this patient a week ago, could you describe how having an earlier referral would have helped you planning this patient's discharge?	Exploring mechanisms to improve discharge outcomes. M1=early referral gives staff more time to assess patients and organised services needed.
13	I have read in another evaluation somewhere else that, since the new programme exists, social services are more thorough in asking for health assessments (like mental health assessments) to gain a bit of time. Could you describe how would that benefit the discharge planning of this patient?	Exploring mechanisms to avoid the fine and unintended outcomes

Box 1. Theory gleaning example for unintended outcome 1 (Patients are discharged without meaningful involvement in decision making). Extract from field notes summarising informal conversations and thoughts before and after realist interview with programme end-user.

'Arrived to the ward on Monday morning. On the board, next to the patient's name says that ambulance for Mrs Leachman has been booked for Tuesday a.m. to be transferred to the local community hospital. I spoke with staff nurse and told her that I was going to speak to Mrs Leachman and she says 'She is going to the community hospital tomorrow'. I said 'Yes, I know'. Visited patient at around 11.30 a.m. Started saying that I have been told that she is going tomorrow, and she answers 'Am I? Good!' I asked if she didn't know anything about it and she says she didn't. Nobody has said anything. I am totally shocked as I found out last Friday morning [three days ago because today is Monday] when I was at the social services office, because the discharge liaison officer in the hospital rang social services to tell them. Also somebody wrote on the ward white board that Mrs Leachman was going tomorrow. But today is Monday and the patient hasn't been informed yet. I told the patient if it's OK I could come back at around 1 p.m. for the interview, they'll probably talk to her about her discharge this morning. She agrees. I went to the nursing station and talked to the nurse in charge and told her that I went to see the patient and she didn't know that she is going tomorrow. She said 'Yes, I know, I haven't had time to speak to her yet'. I explained that I am going to talk about her discharge plans and it'll be good if she had the information of when she is going before I talk to her today at 1 p.m. [never mind just because she needs to know!]. She said yes, she'll talk to her.

I come back to the ward at around 1.15 p.m. Patient sat on her chair (as previously) reading a book. I asked if it is OK to have the interview now and she agrees, although she tells me that nobody has been to talk to her about her discharge yet. I asked if she'd rather me come tomorrow morning and she says she prefers to talk to me now, if she is going to go tomorrow. We have a long chat. I leave at around 2.10 p.m. and nobody has been to talk to her. Nurses may now go into handover and I wondered if anybody will ever talk to her, as it happened in this ward with Case 1 who was informed by the ambulance man that she was going to the rehabilitation home' (Notes from interview with Case Study 12).

References

BENNEY, M. & HUGHES, E. C. 1956. Of sociology and the interview. *American Journal of Sociology,* 62**,** 137-142.

BLAMEY, A. & MACKENZIE, M. 2007. Theories of Change and Realistic Evaluation Peas in a Pod or Apples and Oranges? *Evaluation,* 13**,** 439-455.

BOYCE, C. & NEALE, P. 2006. *Conducting in-depth interviews: A guide for designing and conducting in-depth interviews for evaluation input* Watertown, MA:, Pathfinder International.

BRIGGS, C. 1986. *Learning How to Ask*. Cambridge: Cambridge University Press.

CAMPBELL, D. T. 1975. "Degrees of Freedom" and the case study. *Comparative Political Studies,* 8**,** 178-193.

CHEN, H. T. & ROSSI, P. H. (eds.) 1992. *Using Theory to Improve Program and Policy Evaluations. ,* Santa Barbara, CA: Greenwood Press.

CLARK, A. M. 2008. Critical realism. In L. Given (Ed.), *The SAGE Encyclopedia of Qualitative Research Methods*. pp. 167-170 .Thousand Oaks, CA: Sage Publications.

CORBIN, J. & STRAUSS, A. 2008. *Basics of Qualitative Research: Tecniques and Procedures for Developing Grounded Theory,* London, Sage.

DAHLER-LARSEN, P. 2011. *The evaluation society. ,* Stanford, CA., Stanford University Press.

DALKIN, S. M., GREENHALGH, J., JONES, D., CUNNINGHAM, B., & LHUSSIER, M. 2015. What's in a mechanism? Development of a key concept in realist evaluation. *Implementation Science*, 10(1), 49.

DENZIN, N. K. & LINCOLN, Y. S. 2000. Introduction: The discipline and practice of qualitative research. *In:* DENZIN, N. K. & LINCOLN, Y. S. (eds.) *Handbook of Qualitative Research.* Thousand Oaks, CA: Sage.

DIXON-WOODS, M., SHAW, R. L., AGARWAL, S. & SMITH, J. A. 2004. The problem of appraising qualitative research. *Quality and Safety in Health Care,* 13**,** 223-225.

EMMEL, N. 2013. *Sampling and Choosing Cases in Qualitative Research: A Realist Approach. ,* London, Sage.

ETHERIDGE, F., COUTURIER, Y., DENIS, J. L., TREMBLAY, L. & TANNENBAUM, C. 2013. Explaining the success or failure of quality improvement initiatives in long-term care organizations from a dynamic perspective. *Journal of Applied Gerontology***,** 1-18.

FIELDING, N. & THOMAS, H. 2001. Qualitative interviewing. *In:* GILBERT, N. (ed.) *Researching social life.* London: Sage.

GOFFMAN, E. 1959. *The presentation of self in everyday life,* London, Penguin.

GREENHALGH, T. 1999. Narrative based medicine in an evidence based world. *British Medical Journal,* 318**,** 323-325.

GREENHALGH, T., HUMPHREY, C., HUGHES, J., MACFARLANE, F., BUTLER, C. & PAWSON, R. 2009. How do you modernize a health service? a realist evaluation of whole-scale transformation in London. *Milbank Quarterly,,* 87**,** 391-416.

GUBRIUM, J. F. (ed.) 2012. *The Sage handbook of interview research: The complexity of the craft. ,* London: Sage.

GUEST, G., ARWEN, B. & JOHNSON, L. 2006. How many interviews are enough? An experiment with data saturation and variability. *Field Methods,* 18**,** 59-82.

HAMMERSLEY, M. 1992. Ethnography and realism. In M. Hammersley, *What's wrong with ethnology? Methodological explorations* (pp. 43-56). London, Routledge.

HAMMERSLEY, M. 2008. *Questioning Qualitative Inquiry: Critical Essays*. Thousand Oaks, CA., Sage.

HENRY, G., JULNES, G. & MARK, M. 1998. Realist evaluation: An emerging theory in support of practice. *New Directions in Program Evaluation* 78. San Francisco: Jossey-Bass.

HOLSTEIN, J. A. & GUBRIUM, J. F. 1997. *The Active Interview,* London, Sage.

IOSIFIDES, T. 2011. A generic conceptual model for conducting realist qualitative research. Examples from migration studies. *In:* INSTITUTE, I. M. (ed.) *Working Papers.* Oxford, UK: Oxford Department of International Development (QEH), University of Oxford.

KVALE, S. 1996. *Interviews. An Introduction to Qualitative Research Interviewing,* Thousand Oaks, CA, Sage.

KVALE, S. & BRINKMANN, S. 2009. *InterView: Learning the Craft of Qualitative Research Interviewing,* Thousand Oaks, CA, Sage.

LAWRENCE-LIGHTFOOT, S. 2005. A dialogue between art and science. *Qualitative Inquiry,* 11**,** 3-15.

LINNAN, L., & STECKLER, A. 2002. Process evaluation for public health interventions and research (pp. 1-24). San Francisco, California: Jossey-Bass.

MADILL, A. 2008. Realism. in L. Given (Ed.), *The SAGE Encyclopaedia of Qualitative Research Methods*. pp. 731-735. Thousand Oaks, CA: Sage Publications.

MALUKA, S., KAMUZORA, P. & SANSEBASTIAN, M. 2011. Implementing accountability for reasonableness framework at district level in Tanzania: a realist evaluation. *Implementation Science,* 6**,** 1-15.

MANZANO, A., & PAWSON, R. 2014. Evaluating deceased organ donation: a programme theory approach. *Journal of Health Organization and Management*, 28(3), 366-385.

MANZANO-SANTAELLA, A. 2011. A realistic evaluation of fines for hospital discharges: incorporating the history of programme evaluations in the analysis. *Evaluation*, 17(1), 21-36.

MARCHAL, B., VAN BELLE, S., VAN OLMEN, J., HOEREE, T. & KEGELS, G. 2012. Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research. . *Evaluation,* 18**,** 21.

MARK, M. M., HENRY, G. T. & JULNES, G. 1999. Toward an integrative framework for evaluation practice. *American Journal of Evaluation,* 20**,** 177-198.

MASON, J. 2002. Qualitative interviewing: Asking, listening and interpreting. *Qualitative Research in Action*, 225-241.

MASON, M. 2010. Sample size and saturation in PhD studies using qualitative interviews. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 11.

MAXWELL, J. 2012. *A Realist Approach for Qualitative Research*. Thousand Oaks, CA: Sage.

MAXWELL, J. 2013. *Qualitative Research Design: An Interactive Approach* (3rd edition). Thousand Oaks, CA: Sage Publications.

MISHLER E.G. 1991. *Research Interviewing: Context and Narrative*. (2nd ed). Cambridge, MA: Harvard University Press.

MORSE, J. 1995. Editorial: The significance of saturation. *Qualitative Health Research,* 5(2), 147-149.

OAKLEY, A. 1981. Interviewing women. *In:* ROBERTS, H. (ed.) *Doing Feminist Research.* London: Routledge.

PATTON, M. Q. 1999. Realistic evaluation [Review of the book Realistic Evaluation]. *American Journal of Evaluation.,* 20**,** 385-388.

PATTON, M. Q. 2002. *Qualitative Research and Evaluation Methods,* Thousand Oaks, CA., Sage.

PATTON, M. Q. 2003. Qualitative evaluation checklist. Evaluation Checklists Project.

PAWSON, R. 1989. *A Measure for Measures. A Manifesto for Empirical Sociology,* London, Routledge.

PAWSON, R. 1996. Theorizing the interview. *The British Journal of Sociology,* 42**,** 20.

PAWSON, R. 2006. Digging for nuggets: How 'bad'research can yield 'good'evidence. *International Journal of Social Research Methodology,* 9**,** 127-142.

PAWSON, R. 2013. *The Science of Evaluation: A Realist Manifesto.* London, Sage.

PAWSON, R. & MANZANO-SANTAELLA, A. 2012. A realist diagnostic workshop. Evaluation, 18(2), 176-191. *Evaluation,* 18**,** 16.

PAWSON, R. & TILLEY, N. 1997. *Realistic Evaluation,* London, Sage.

PAWSON, R. & TILLEY, N. 2004. Realistic Evaluation. *In:* OFFICE, B. C. (ed.).

POPPER, K. 2002. *The Logic of Scientific Discovery*. London: Routledge.

RUBIN, H. J. & RUBIN, I. S. 1995. *Qualitative Interviewing. The Art of Hearing Data,* Thousand Oaks, CA., Sage.

RYCROFT-MALONE, J., FONTENLA, M., BICK, D. & SEERS, K. 2008. Protocol-based care: impact on roles and service delivery. *Journal of Evaluation in Clinical Practice,* 14**,** 867-873.

SAPSFORD, R. & ABBOTT, P. 1992. *Research Methods for Nurses and the Caring Professions,* Buckingham, Open University Press.

SAUMORE, K, and GIVEN, L. 2008. Data saturation. pp. 195-196 in L. Given (Ed.), *The SAGE Encyclopaedia of Qualitative Research Methods.* Thousand Oaks, CA: Sage Publications.

SAYER, A. 1992. *Method in Social Science: a Realist Approach.,* London, Routledge.

SEIDMAN, I. 2012. *Interviewing as Qualitative Research: A Guide for Researchers in Education and the Social Sciences,* New York, Teachers college press.

STAKE, R. E. 2010. *Qualitative research: Studying how things work. ,* New York, Guilford Press.

STRAUSS, A., FAGERHAUGH, S., SUCZEK, B., et al., 1985. *Social Organization of Medical Work*. Chicago, The University Of Chicago Press.

TILLEY, N. 2000. The evaluation jungle. *In:* BALLINTYNE, S. P., KENNETH; MCLAREN, VIC (ed.) *Secure Foundations: Key Issues in Crime Prevention, Crime Reduction and Community Safety* London: Institute for Public Policy Research.

TILLEY, N. 2009. What's the 'what' in 'what works'? Health, policing and crime prevention. *In:* KNUTSSON, J. T., N, (ed.) *Evaluating Crime Reduction Initiatives.* Monsey, NY: Criminal Justice Press.

WEISS, C. H. 1998. *Evaluation: Methods for Studying Programs and Policies. ,* London, Prentice-Hall.

WESTHORP, G. 2013. Developing complexity-consistent theory in a realist investigation. *Evaluation*, 19(4), 364-382.

YIN, R. K. 2010. *Qualitative Research from Start to Finish,* New York, Guildford Press.